

DOCUMENT RESUME

ED 193 261

TM 800 575

AUTHOR Becker, Marc S.; Reed, Larry
 TITLE Testing for Accountability: A Practical Guide for Administrators and Teachers.
 INSTITUTION Arizona State Univ., Tempe. Coll. of Education.
 PUB DATE May 80
 NOTE 89p.
 AVAILABLE FROM Arizona Educational Information System, Farmer Bldg., Room 105, Arizona State University, Tempe, AZ 85281 (\$4.50).

EDRS PRICE MF01/PC04 Plus Postage.
 DESCRIPTORS *Accountability; *Criterion Referenced Tests: Elementary Secondary Education: Mathematics; *Norm Referenced Tests: Program Implementation: Public Education: Reading Achievement: Scoring: State School District Relationship; *Testing Programs: Test Interpretation; *Test Selection
 IDENTIFIERS Arizona

ABSTRACT

This resource guide was designed for teachers and administrators in Arizona school districts to assist them in preparing functional and meaningful programs of student achievement testing. The format of the guide is based on fundamental questions often posed by educators regarding norm-referenced and criterion-referenced testing. Particular emphasis is placed upon practical guidelines for implementing both types of testing in terms of accountability. The resource guide is divided into two sections. The first section discusses norm-referenced testing, and is applicable to educators desiring this type of test in addition to those mandated by Arizona's accountability legislation. The second section discusses criterion-referenced tests, and provides a rationale for their use. Terminology is nontechnical, information is presented in a readily accessible format, and resources/references are listed. (Author/GK)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *



U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED193261

PREPARED FOR ARIZONA EDUCATORS THROUGH
ARIZONA EDUCATIONAL INFORMATION SYSTEM

TESTING FOR ACCOUNTABILITY

A PRACTICAL GUIDE FOR
ADMINISTRATORS AND TEACHERS



"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

J. Drake

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

MAY 1980

OFFICE of FIELD SERVICES
COLLEGE OF EDUCATION
ARIZONA STATE UNIVERSITY

RESOURCE GUIDE

TM 800575

TESTING FOR ACCOUNTABILITY
A PRACTICAL GUIDE FOR ADMINISTRATORS
AND TEACHERS

Written by:

MARC S. BECKER, ASSISTANT DIRECTOR
Educational Management System
Glendale Union High School District
Glendale, Arizona

LARRY REED, DIRECTOR
Instructional Services
Osborn Elementary School District
Phoenix, Arizona

No portion of this Resource Guide may be produced
without the permission of the Office of Field Services,
College of Education, Arizona State University.

TESTING FOR ACCOUNTABILITY
A PRACTICAL GUIDE FOR ADMINISTRATORS
AND TEACHERS

Prepared by: Arizona Educational Information System

AEIS STAFF

Dr. Ronald J. Madsen
Director

Mrs. MaryLee Carter
Program Coordinator

Ms. Nena Ward
Information Specialist

Mrs. Mary Gabrys
Secretary

AEIS BOARD OF DIRECTORS

Dr. Donald Wright, Chairman,
Superintendent, Phoenix Elementary
District

Dr. Paul Plath, Vice-Chairman,
Superintendent, Yuma UHS District

Dr. Leonard Cahen, Member, Associate
Dean, College of Education - ASU

Dr. Lou Carey, Member, Director,
Office of Research Services,
College of Education - ASU

Dr. William Dabb, Member,
Superintendent, Cartwright Elementary
District

Dr. Jackson M. Drake, Member,
Director, Office of Field Services,
College of Education - ASU

Dr. William Greaves, Member,
Superintendent, Osborn Elementary Dist.

Dr. Pat Henderson, Member,
Superintendent, Phoenix UHS District

Dr. Raymond Kellis, Member
Superintendent, Peoria Unified District

Dr. E. T. Luty, Member
Superintendent, Globe Unified District

Mr. Archie Stephens, Member
Superintendent, Humboldt Unified Dist.

PREFACE

The purpose of education is to create an informed citizenry. We professional educators have been entrusted with the task of preparing and implementing appropriate instructional programs for students in the local school district in order to accomplish this purpose. It is not an easy job. The public holds us accountable, and expects us to verify the excellence of our work with some form of evaluation.

Testing is the vanguard of the public school instructional program, and should be an integral part of that program. Appropriate and humane decisions about curriculum and instruction are enhanced when teachers and administrators have solid information about student achievement. Local taxpayers, parents, and the Board of Trustees have the right to hold the schools accountable for educating their youth. In Arizona, the burden of proof of student achievement has been placed on the local school district by a mandate from the State Legislature. House Bill 2013 makes it obligatory for school districts to adopt and begin to carry out a program of student achievement testing by January, 1981. Recognizing that this legislation is a reality, and that school districts will need resources to assist them in preparing functional and meaningful programs, AEIS has prepared this resource guide entitled Testing For Accountability: A Practical Guide for Administrators and Teachers.

We appreciate the careful attention that the authors, Mr. Larry Reed and Mr. Marc Becker, have given to researching the topic and the time they have accorded to preparing the manuscript. It is with confident pride that we share this monograph with you.

Dr. William F. Greaves, Superintendent
Osborn Elementary District
Member, Board of Directors
Arizona Educational Information
System

FOREWORD

The advent of criterion-referenced tests in the last decade has complicated the issue of competency-based testing. Educators now have an alternative means of evaluating student achievement besides the norm-referenced test that has been traditionally used in school districts across the country. The authors of this resource guide do not attempt to engage the support of the reader for the use of one test in preference for another. Instead, they discuss the specific purpose of each of these alternatives with regard to determining student achievement and assessing accountability as complements to each other.

Those who support accountability have not always agreed on the question of which type of testing to employ and the authors have carefully outlined, compared, and contrasted criterion and norm-referenced testing to point out the beneficial uses of each type of test instrument. They emphasize that effective testing programs need clearly defined goals and objectives that are compatible with the curriculum-instruction goals and objectives of each local school district.

The format of this guide is based on fundamental questions often posed by educators and which the authors have encountered through the development and implementation of testing programs in their districts. It is our intention that Testing For Accountability: A Practical Guide For Administrators And Teachers, will provide useful information on both alternative means of testing for accountability, and will be of assistance to those school districts whose testing programs are used to make such an assessment of the effectiveness of our educational system.

Dr. Ronald J. Madsen, Director
Arizona Educational Information
System

ABOUT THE AUTHORS

Larry Reed, author of the norm-referenced testing section of the resource guide, has been a resident of Arizona since 1938. He is a graduate of Whittier Elementary School, North Phoenix High School, Phoenix College, and received his B. A. in Education and M. A.'s in Education, and Administration and Supervision from Arizona State University.

Mr. Reed taught for 11 years in middle and junior high grades, was principal of Longview School (grades K-8), and served as principal of Grandview School (grades 5-8). Currently Mr. Reed is the Administrator for Instructional Services of the Osborn School District #8 in Phoenix.

Marc Becker, who authored the criterion-referenced testing section of the resource guide, received his B. A. in Psychology from California State University, an M. S. in Adult Education from Cornell University in New York, and his Ph. D. in Educational Research and Development from Florida State University.

Mr. Becker taught English at Eli Cohen Elementary School in Israel, implemented courses in Adult Education as a Teaching Assistant at Cornell University, served as a Research Assistant in the Center for Educational Design and Evaluation, the Career Education Center, and the Office of Dissemination and Diffusion at Florida State University. Mr. Becker is currently the Assistant Director of the Educational Management System for the Glendale Union High School District in Glendale, Arizona.

ACKNOWLEDGMENTS

We gratefully acknowledge the following Arizona State University personnel for their direction and support.

MaryLee Carter, monograph project director, editor

Margaret Embury, editor

Donna Chapman, manuscript preparation

Mary Gabrys, manuscript preparation

Nena Ward, manuscript preparation

Janet Berg, reproduction

TABLE OF CONTENTS

	Page
PREFACE	iii
FOREWARD	iv
ABOUT THE AUTHORS	v
 Chapter	
I. INTRODUCTION	1
II. THE IMPLEMENTATION OF NORM-REFERENCED TESTING IN SCHOOLS By Mr. Larry Reed	4
a. When Should You Test?	7
b. What Should You Test?	9
c. How Do You Choose a Test?	10
d. How Do You Shop for a Norm-Referenced Test?	15
e. What Scoring Services are Available?	16
f. What Scores are Used to Interpret Norm-Referenced Tests?	18
g. Who is Responsible for Test Program Management and Data Dissemination?	29
h. References	32
III. THE IMPLEMENTATION OF CRITERION-REFERENCED TESTING IN SCHOOLS By Mr. Marc Becker	33
a. What is a Criterion-Referenced Test?	35
b. How do Criterion-Referenced Tests Differ from Norm-Referenced Tests?	38
c. Why use Criterion-Referenced Tests?	41
d. What does a Criterion-Referenced Test Program Consist of?	45
e. What are the Steps Involved in Setting Up a Criterion-Referenced Testing Program?	47
f. References	62

	Page
IV. APPENDICES	63
Appendix A Glossary	63
Appendix B House Bill 2013	65
Appendix C Bibliography of AEIS Documents	67

Within the last decade, testing has assumed an importance unparalleled in the history of American education. In every corner of the nation, taxpayers and their elected representatives are looking closely at test results to determine whether schools are educating students effectively. This increased public interest in testing comes in the wake of the sharp decline in student achievement which characterized American education in the 1960's and the 1970's. Educators must now face a new reality--testing is inextricably linked to accountability. In Arizona, as elsewhere, legislation mandates that schools undertake new efforts to improve student learning. Testing is the vehicle by which the results of these efforts are communicated to the public. It is also a tool for making decisions about the placement or remediation of students and the improvement of instruction.

Educational policymakers who are charged with the implementation of accountability face an important question: What type of testing should be used to measure student achievement? A few years ago, this question would not have been asked. Only one type of testing was available--the norm-referenced type in which student achievement is compared with national norms. Within the last decade, an alternative approach to testing has emerged across the nation. This approach to testing compares student achievement with well-defined standards. It is known as "criterion-referenced testing".

The purpose of this resource guide is to provide administrators and teachers in local school districts with some basic

information about norm-referenced and criterion-referenced testing. Particular emphasis is placed upon practical guidelines for implementing both types of testing in the service of accountability. No attempt is made to enter the current controversy concerning the merits of norm-referenced vs. criterion-referenced testing. The assumption is made that both types of testing serve valid purposes, and that they can coexist in a district committed to the implementation of accountability.

The resource guide is divided into two sections. The first section, authored by Larry Reed of the Osborn Elementary School District, discusses norm-referenced testing and provides some guidelines for its implementation. This section will be useful to educators who are considering the use of norm-referenced tests in addition to those which are mandated by Arizona's accountability legislation. The second section of the resource guide is authored by Marc Becker of the Glendale Union High School District. This section discusses criterion-referenced testing and provides some steps for its implementation. A rationale is provided for the use of criterion-referenced tests in addition to the norm-referenced tests already being administered in most districts.

Both portions of the resource guide are written for the busy educator who has little time to become acquainted with the technology of testing. The terminology employed is nontechnical, and information is presented in a format which makes it readily accessible. Resources and references are listed in order to assist administrators and teachers in receiving information and in obtaining the services required for test implementation.

The authors of this resource guide sincerely hope that administrators and teachers throughout Arizona will find it useful in making decisions about the implementation of testing.

Larry Reed, Director
Instructional Services
Osborn Elementary District
Phoenix, Arizona

Marc S. Becker, Ass't Director
Educational Management System
Glendale Union High School
District
Glendale, Arizona

THE IMPLEMENTATION OF
NORM-REFERENCED TESTING IN SCHOOLS

By Larry Reed, Director
Instructional Services
Osborn Elementary School District
Phoenix, Arizona

Norm-referenced testing is one of the most efficient methods of securing information that is valid, reliable and relevant for a broad spectrum of objectives.

It is also generally part of the testing program mandated by most states' legislatures in an attempt to evaluate the educational program within its area of control. Many legislators are accustomed to using norm-referenced testing as a means of evaluating student achievement because it provides a normed measurement that is easily interpreted against a national average and clearly reflects gain or loss over an extended period of time.

To view norm-referenced testing as only an instrument to satisfy state mandate is a gross waste of a valuable information gathering source.

A comprehensive testing program includes more than just standardized tests. In addition to standardized tests, it should include other components (e.g., competency-based testing (criterion-referenced) for establishing basic skills requirements; diagnostic testing (normed or criterion-referenced) to assist in organization for instruction for small groups or individual students; teacher-made tests to assess specific objectives taught in a given course or unit of instruction; and of course, a norm-referenced test that can provide a method of evaluation for large groups and total program effectiveness in the areas tested).

When properly used the standardized test can provide valuable data for the following purposes:

1. Diagnosis of individual instructional needs. The use of item analysis is best applied for this purpose.

2. A logical and understandable basis for promotion and retention of students. These data coupled with academic standing and teacher recommendation tend to provide a defensible and, to some degree, acceptable criterion for student promotion or retention.

3. Needs assessment information to qualify for Federal program funding and assistance.

4. To assist in the grouping of students for instructional purposes i.e., regrouping in certain content areas, homogeneous grouping for class assignment such as gifted or remedial classes, tracking, leveling or any other means of bringing students together with common instructional needs or learning patterns.

5. To evaluate student progress after a reasonable period of instruction in the objectives tested.

6. For instructional and institutional decision making regarding course content, time allotments and instructional materials.

7. For the evaluation of instructional effectiveness. In order for this to have any validity, data about student academic aptitude must be available.

CAUTION

If the results of norm-referenced tests are to be used as a measure of instructional or program effectiveness there must be an accompanying measure of student potential.

Comparison to national norms without consideration of the ability of the students taking the test cannot provide truly valid data for assessment.

8. To compare student achievement to a large normed population.

9. To predict academic course offerings in the total instructional program.

10. To assist with personal, educational and vocational planning.

Norm-reference test data can be a useful tool to students, parents, teachers and administrators in making those judgments and decisions necessary to provide an instructional program designed to meet the needs of all students.

A norm-referenced testing program is a vital part of the checks and balance system needed to maintain a quality instructional program in any school district.

The norm-referenced test should never become the sole determinate of the curriculum, nor should it be the only criterion for measuring of student achievement, teacher competency or program content. Testing is a device to aid instruction and learning, not the means to attain them.

* * * * *

A norm-referenced test is designed to compare students with national norms and to predict trends.

It is a tool to classify students into categories of achievement and to measure achievement in broad areas of knowledge. Norm-referenced tests are only a part of a total testing program and should never be the only instruments used to evaluate program effectiveness.

WHEN SHOULD YOU TEST?

The reasons for which you test will determine when you test.

1. Fall-only testing. It is recommended that this be done during the second or third week of school. The prime purpose of testing only in the fall is to provide data to assist with class placement (grouping), and to diagnose student needs (using item analysis) for organization and instruction. Hand scoring of the test would be necessary in order to have data available for immediate use.
2. Mid-Year-only testing. It is recommended that this be done prior to the end of the first academic semester. The data from mid-year testing are valuable for curriculum revision if deficiencies appear, regrouping of students for instructional purposes, evaluation of student progress as compared to national norms, consideration of pass/fail for some students, and to assess teacher effectiveness providing some measure of student potential is available for correlation.
3. Spring-only testing. If testing is to be done only once a year the greatest value may be obtained from the spring-only test. It is recommended that this test be administered the first or second week of May. Testing at this time presents the greatest value for making administrative decisions about curriculum, and the least value for student diagnosis. It will provide data to measure student achievement as compared to local and national norms, and a measure of student progress if data from previous years are available. It can supply information necessary to conduct a needs assessment for curriculum reorganization and for qualifying for Federally funded programs.

It can be useful in reaching and defending decisions for promotion or retention of students. In this instance, hand scoring would be required due to lengthy turn-around time required where using a computer scoring service. It can be useful in establishing class needs and student grouping for the coming year.

It is an excellent tool for the evaluation of strengths and weaknesses of program and instructional materials.

Finally, it can provide useful data for evaluation of teaching effectiveness when compared to student academic aptitude.

4. Fall-and-Spring testing. Ideally, this is the best testing program. By pre- and posttesting it provides everything all the single test schedules do, as well as meets the needs for evaluation required for most Federally funded programs. When used for Federal program evaluation, careful attention should be paid to the empirical norming dates of the instrument. Most

programs require that the test be administered two weeks before or after these dates.

As in the case of the spring-only test, the administration has additional time during the non-school months to analyze the test data for predicting and analyzing curriculum needs and strengths, and to consider additional methods of evaluation prior to making changes.

WHAT SHOULD YOU TEST?

The majority of norm-referenced tests are developed to measure the critical "core" subject areas (e.g., phonic and structural analysis; reading vocabulary; reading comprehension; language usage; grammar; spelling; arithmetic fundamentals; concepts and applications; and in some basic packages the reference skills are included). Many also have "add-ons" for social studies and science.

Some publishers offer a compatible test for measuring student ability or academic potential that is closely correlated to the various sub-sections of the test battery.

Obviously, norm-referenced testing is not intended to be a total evaluation plan. Many areas must be measured by using other types of instruments, primarily criterion-referenced tests, either locally prepared or purchased from various publishers.

REMEMBER

Norm-referenced tests are not designed to be totally compatible with a specific instructional program. They usually measure student achievement in broad areas of knowledge rather than specific skills. They provide a basis for investigation of programs, but lack the precision of a criterion-referenced test for critical evaluation.

HOW DO YOU CHOOSE A TEST?

There are many important points to consider in selecting a norm-referenced testing program. The following criteria are recommended for reviewing the batteries prior to selection.

1. Compatibility with curriculum. Under no circumstances should any test determine the curriculum. The curriculum content should dictate the test selected. Analysis of the test objectives provided by the publisher and the test items themselves as compared to course content, scope and sequence, and specific instructional material should provide an excellent indicator to determine if the test is compatible with the curriculum. Special attention should be paid to the instrument's ability to test any district or state basic skills mandates.

2. K-8 continuity. Data should be available to show between level reliability. A given derived score, such as grade equivalent on one level, should be a predictor of an individual score on contiguous levels.

Another consideration is the number of test levels and forms available. Ideally, there should be a different test for each grade level, and at least two forms for each level.

3. K-12 continuity. The high school extension of the test battery should match the K-8 program. In a unified district this would occur automatically, but in situations where the elementary is separate from the secondary organization, it would be advisable to attempt coordination of the program.

4. Measure of academic potential. The total program should include an instrument correlated and compatible with the academic battery that could provide data for comparison of student potential with academic achievement attained. This instrument should have the same technical excellence as its companion test.

5. Technical excellence. The technical manual should indicate excellence as shown by validity, reliability, range coverage and score gradation.

6. Minimum test bias. Careful analysis of the battery should be made to determine the degree of sex bias and cultural bias contained in the test.

7. Relationship to total local testing program. Skills measured by norm-referenced tests should be compatible to other elements of the total testing program (e.g., criterion-referenced tests and teacher-made tests).

8. Equal subject area coverage. All subject areas in the battery should be equally valid and weighted. This generally becomes evident when doing the item match with the objectives provided by the publisher and the curriculum.

9. Specialized norms. It is important that norms be representative of the population to be tested.

10. Equal level ratings. It should be determined if all levels of the battery are equally sound and valid. Do not limit test scrutiny to any one level. Look carefully at them all.

11. Format. The test format should be appropriate to the students' level of understanding. It should not overwhelm the student with content and decisions. There should be a minimum of visual and verbal distractors.

12. Valid out-of-level testing. Valid norms should be available for students taking tests at levels substantially below their grade placement. It is important to consider what the effect out-of-level testing might have on the validity of the reporting program's strengths and weaknesses, and/or other statistical reporting.

13. Placement tests. If out-of-level testing is to be considered, there should be locator/diagnostic tests available to determine the appropriate test level for the student.

14. Practice tests. If a test is to measure knowledge, not ability to take a test, then appropriate practice instruments should be available at all levels. This, coupled with your content-item-objective match, can help to insure instructional validity.

15. Test administration. Examiner's manuals should provide simple, concise, explicit and easy to follow directions for the teachers.

16. Timing. The time required for test taking should be reasonable for the age group being tested. The test's sub-sections should be segmented to allow for variations in attention span of the students, and to allow for flexibility in dividing the testing task.

17. Scoring services. The effectiveness of any testing program may depend on the quality, availability and flexibility of the scoring package. Items for consideration are:

- a. Ability to hand score if necessary. This includes all the necessary scoring tables for raw score conversion to required reporting format.

b. The flexibility of the computer scoring service offered by the publisher to provide the following:

- . individual test records
- . class record sheets
- . stick-on labels for student records
- . graphic reporting system for parents, students and teachers
- . administrator's summary
- . local norms
- . frequency distributions
- . school and district item summary
- . evaluator's summary
- . pre-post evaluator's summary
- . summary report on program strengths and weaknesses.

c. The ability to adapt the scoring program to a school district's computer.

d. The information given for dissemination and evaluation for accountability (e.g., teacher data, parent/student print out, etc.).

18. Cost. Nothing can be done without consideration of cost and budget. The old adage "you get what you pay for" was never more correct than in the selection of a testing program.

Generally, the following items or services determine the cost:

- . test booklets for total battery and for aptitude test (Examiner's manuals are usually free)
- . consumable tests for primary levels
- . answer sheets
 - 1. hand score or self scoring
 - 2. computer bubble sheets
- . practice tests

- . class management guides
- . norm books
- . scoring services
- . shipping charges.

19. Other optional services and/or materials. Many test publishers offer supportive programs and materials other than those previously mentioned such as:

- . diagnostic and/or criterion-referenced tests that are correlated with the norm-referenced test program
- . specific instructional materials keyed to the tests used
- . home activity materials for parents and students
- . parent's guide for understanding tests and test results
- . individual student status reports
- . resource guides to assist teachers in applying test results to special curriculum materials.

A rough estimate of cost for a basic once-a-year testing program is approximately \$1.50 per child for tests and supportive materials, and approximately \$2.00 per child if the program includes complete computer scoring services. This, of course, would vary depending upon the individual district's needs and budget considerations.

A small school district might reduce the original costs of its testing program by entering into a "coop" with one or two other small districts, and parceling out the purchase of non-consumable items. This would entail a central housing and distribution system, and close communication in establishing testing dates.

A list of criteria similar to these is used by the Atlanta

Public Schools, and to some extent, by the members of the Greater Phoenix Curriculum Council, in selecting their norm-referenced testing programs. Each of the criterion was placed on a decision grid on which all batteries under consideration were rated. Each criterion was rated on a scale of one to five of which five was considered the highest. A representative committee was formed to evaluate the test batteries, and the grid format was used to direct questions to the various company representatives, and used in analysis of the testing programs.

* * * * *

The only single factor that led to rejection of any battery of tests was its incompatibility with the curriculum. Therefore, the prime consideration in judging any test battery is the assessment of how harmonious it is with the curriculum.

4

HOW DO YOU SHOP FOR A NORM-REFERENCED TEST?

Once the selection team has been named, and the criteria established for test evaluation, the next step is to make contact with recognized and established test publishers to arrange for examination of their materials, and to hear formal presentations by their test representatives.

The following is a partial list of the major publishing companies who market norm-referenced tests:

Addison Wesley
2725 Sand Hill Rd
Menlo Park, CA 94025

Riverside Publishing Company
1919 South Highland Ave
Lombard, IL 60148

CTB McGraw Hill, Inc.
1700 E. Don Carlos #1-U
Tempe, AZ 85281
Phone 602 894-1926

Science Research Associates (SRA)
155 N. Wacker Drive
Chicago, IL 60606

Harcourt Brace, Jovanovich
Test Dept.
757 Third Ave
New York, NY 10017

Scott Foresman
Test Dept.
855 California Ave
Palo Alto, CA 94304

The Psychological Corp
757 Third Ave
New York, NY 10017

WHAT SCORING SERVICES ARE AVAILABLE?

Determining the scoring service to be used is as essential to the testing program as the selection of the test itself. The three basic options available are hand scoring, central scoring within a given school system that has the capacity to do so (computer) and contracting with the test publisher or a commercial agency to perform the service.

	Advantages	Disadvantages
Hand Scoring	1. immediate feedback of data 2. low cost 3. no time lost for reporting.	1. loss of instruction time for teacher scoring 2. limitations on data available from test 3. margin of error greater 4. requires extensive clerical time 5. all recording and reporting must be done by hand.
Central Scoring	1. lower cost (if system exists) 2. shorter "turn-around" time for reporting 3. ability to customize reporting 4. minimum teacher time 5. ability to include other test data in reports 6. home storage of data.	1. requires computer 2. requires appropriate programming 3. requires extensive personnel 4. requires maintenance of forms for recording and reporting 5. overall very expensive to originate.

	Advantages	Disadvantages
Contract Scoring	1. no additional personnel required 2. no teacher time required other than testing 3. almost unlimited data available 4. any size district can obtain extensive data for evaluation, program analysis and reporting 5. is efficient and accurate.	1. cost 2. turn-around time from 2 to 4 weeks 3. no local storage of data.

Central scoring would only be feasible in a very large school system due to its complexity of origin and operation.

In general, contract scoring is the most logical for most districts. Even though cost is listed as a disadvantage, in the long run, the cost per pupil is minimal when considered against the value of the service. It permits the very smallest of districts to have as sophisticated a scoring analysis and reporting system as the very largest district.

The components of the scoring service can be as flexible as desired, and be tailored to provide the necessary data needed in an individual district or school system.

WHAT SCORES ARE USED TO INTERPRET NORM-REFERENCED TESTS

The scores to be used in the interpretation of the test results depends on the purposes for testing, and the uses to be made of the results.

Various types of scores can be provided by the test publishers. The most common of these are:

Raw scores (RS). The raw score usually is the number of items answered correctly, and is the most basic type of score from which are derived scaled scores, percentile ranks, stanines, grade equivalents and anticipated achievement scores. It can also be the basis for some statistical analysis such as computing coefficients of correlation between scores on two tests. The raw score itself is of little value except in grouping or regrouping small groups of students for instructional purposes.

Grade equivalents (GE). A grade equivalent indicates the grade placement of students for whom a given score is average or typical in specific subject or content areas. It is possible to establish levels of performance to be expected at each grade level. For example, on a one-hundred-item test in reading comprehension, a given raw score of fifty might be typical of an average beginning sixth grade student, and a raw score of seventy, typical of a beginning seventh grade student. In this manner it is statistically possible to convert raw scores to grade equivalent scores (grade plus number of months in that grade).

Date of Testing:	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June
Grade Placement:	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9

The grade equivalent score is probably the most popular and commonly used method of reporting test results because of its simplicity, and because it is easily understood by all. However, it has begun to lose some of that popularity due to some problems in interpretation.

One problem is that extreme scores are readily misinterpreted. Extreme scores, no matter what the type of score used, tend to be less reliable than scores in the middle range. The further from the average, the less accurate the measure. Grade equivalent units are not equal throughout the scale. An increase in raw score of 10 points may amount to only three months of change in one part of the scale and three years in another part of the scale.

A second problem is that grade equivalents are not directly comparable from one test to another for students who are at the extremes. For example, a student tested at the beginning of grade four who obtains a GE of 6.8 in mathematics and 7.2 in reading comprehension, might actually be doing better in math than in reading in relation to other grade four students. This can be noted in the norm tables for any test that gives both GE and percentile ranks.

Another quite common misinterpretation is that grade equivalent scores represent some kind of standard. Anyone who says,

"I want every child in the state to be reading above his or her grade level", probably does not understand what he is asking. A grade equivalent score indicates typical performance - not necessarily good or bad performance. If all students could be raised in that achievement level, then the grade equivalent norms would have to be raised also, so that they could continue to indicate typical performance.

Finally, laypersons commonly think that a student ought to be in the actual grade corresponding to the grade equivalent obtained on an achievement test. They do not recognize GE as a comparison to typical performance.

These objections to grade equivalent are not intended to convince you not to use them in reporting, but rather to assist you in interpreting these scores to laypersons.

Percentile ranks (file rank). A percentile rank tells the percentage of students whose scores fall at or below a given score. Percentile ranks range from a low of 1 to a high of 99, with 50 always indicating average or typical performance. It indicates the relative standing of a student in comparison with the norm group.

Two "pluses" for the use of percentile rank, and in explaining test results to students and parents, are that they are comparable from one test to another, and they are fairly easy to understand.

One negative factor to keep in mind when using percentile ranks is that they tend to bunch up in the middle of the scale and spread out toward the extremes.

Note Figure 1 that demonstrates how an increase of a few raw score points might translate into an increase of 20 percentile rank points in the middle of the scale and only 2 points at either end.

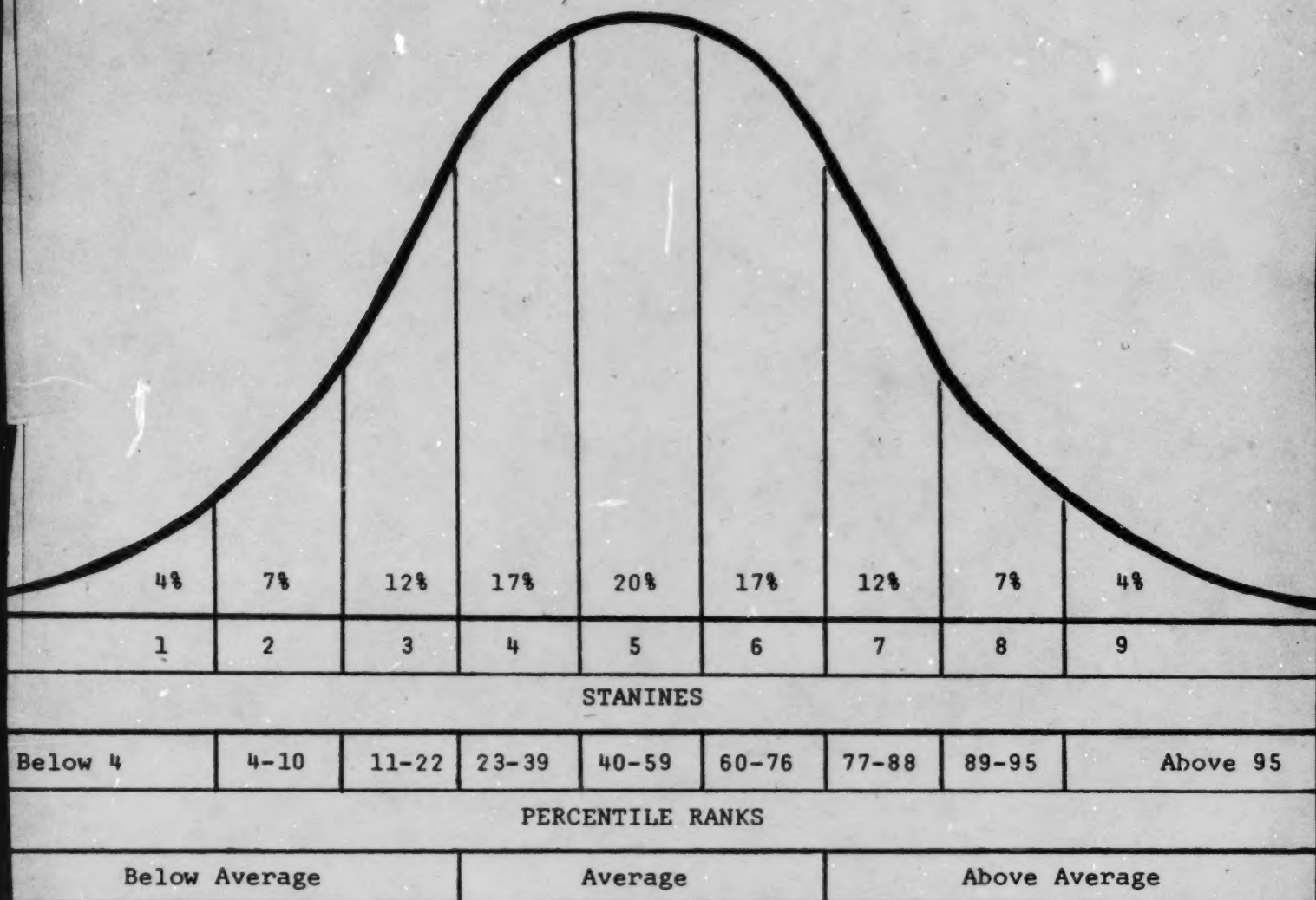


Figure 1 - BROAD CLASSIFICATIONS

2

Stanines. The stanine draws its name from the fact that it is a STandard score from a scale of NINE units. The stanine mean is 5 and the standard deviation is 2. The scale range is from 1 through 9. The intervals between units are equal, making stanines useful for computing means, correlation coefficients, and other statistics. Each unit represents one-half of a standard deviation of the scores in the distribution of the norming group.

Stanines do not "bunch up" or spread out at different points along the scale as percentile ranks and grade equivalents do.

In order for stanines (and percentile ranks) to be accurate, testing should take place at about the time of year that the tests were normed or standardized, roughly speaking about two weeks before or after the empirical norming date. If testing takes place earlier than the time for which the norms being used were determined, the percentile ranks and stanines will be artificially low, i.e., lower than actual student proficiency. If testing is done later than the standard month, results will be artificially high, i.e., higher than actual student proficiency. It is for this reason most Federal programs require that evaluation testing be done within two weeks of the empirical norming data of the instrument used.

It may also be useful to note that when qualifications for certain accelerated programs are established, they are often stated as "a student must be two standard deviations above the norm." Since each stanine unit is one-half of a standard deviation, the student must rank in the ninth stanine to qualify.

When stanines are used for reporting test results to persons

not trained in testing, the following descriptions are appropriate:

1. Stanines 7, 8 and 9 - above average. Students whose scores fall consistently within these levels should be given enriched or accelerated instructional experiences.
2. Stanines 4, 5 and 6 - average. Students scoring consistently in this range can benefit from instruction appropriate to their grade.
3. Stanines 1, 2 and 3 - below average. Students scoring consistently in this category are those for whom a specialized curriculum is desirable.

More specifically:

<u>Stanine</u>	<u>Level</u>
9	Highest level (top 4%)
8	High level (next lower 7%)
7	Well above average (next lower 12%)
6	Slightly above average (next lower 17%)
5	Average (middle 20%)
4	Slightly below average (next lower 17%)
3	Well below average (next lower 12%)
2	Low level (next lower 7%)
1	Lowest level (bottom 4%)

The use of stanines has several advantages. Since each stanine represents a broad band of performance, it has greater stability than other derived scores, and it is often reported reflecting one standard deviation i.e., actual score in the fifth stanine but range reflected on the reporting device would show a band covering stanines four, five and six. For this reason it is a convenient method of communicating test results

to students and parents without the likelihood of misinterpretation or overgeneralization from small, unreliable differences in scores. The ability to express a score as a single digit or band is an advantage, however, this is gained at the expense of precision i.e., not pinpointed on a given scale but a range on a given scale.

Scaled scores (SS) or reference scale scores (RSS). The scale score is a standard score based on a single interval scale extending across all grades and all levels of a test. Scores on the scale are generally expressed in three-digit numbers with a possible range from 001 to 999. (See Figure 2). Information on the development of the scale for a given test is available in the Technical Manual prepared for tests.

Scale scores have little interpretive value, but do perform certain important functions:

1. Scores can be added, subtracted and averaged regardless of the level of the test. Because the differences between two successive scores on the scale is the same over all parts of the scale, differences can be qualified.
2. Direct comparisons among grades are possible, giving a representation of growth rate, because scores on differing levels of a test are related to a single scale rather than to a scale which is specific to a particular grade and time of year.
3. Data are provided that can be used in longitudinal studies of grades or individual students because scale scores are independent of level, grade, time of year and restandardization.
4. The use of test data provides a flexibility for research not anticipated at the time of testing.
5. Provides data that is advantageous for research.

For these reasons scaled scores are suitable for measuring change in achievement in a particular skill or content area over a period of time.

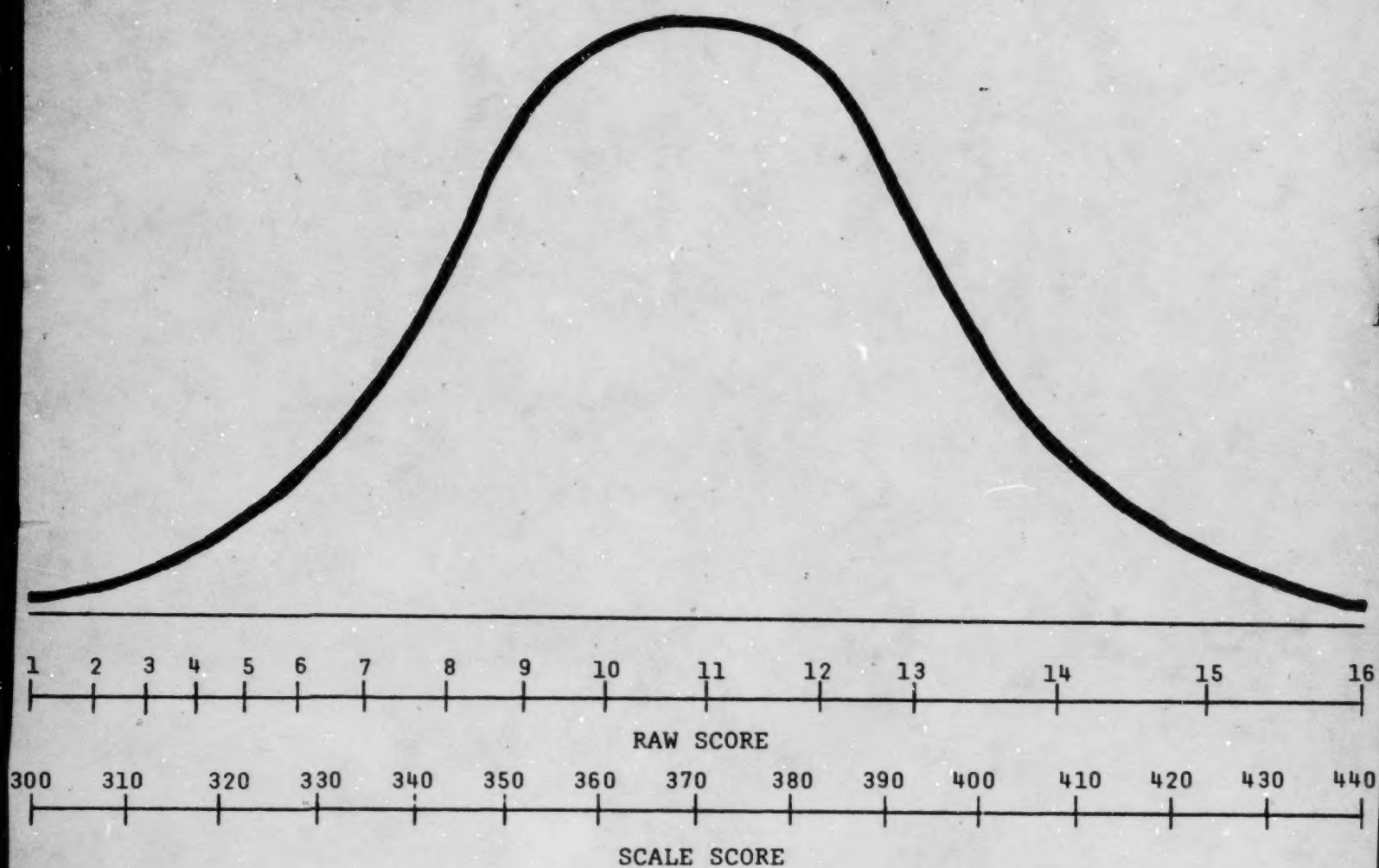


Figure 2 - REPRESENTATION OF THE CORRESPONDENCE BETWEEN
RAW SCORES AND SCALE SCORES

A disadvantage of scaled scores to be noted is that they are not directly comparable from one test to another.

Anticipated achievement scores (AA). The anticipated achievement score is a function of performance on a measure of academic aptitude test correlated to the achievement test. Generally two scores are available: the anticipated achievement grade equivalent and the anticipated achievement scale score (often reflected as an IQ score). The scores are computed for each student for each test, each content area and the total battery by using multiple regression formulas that include as predictors age, grade, sex and the scores of each test.

The test is administered in combination with an achievement test to which it is correlated.

The obvious desirability of such a score is that it provides a measure of student achievement against his/her own potential as well as the norm comparison.

SUMMARY - USE OF SCORES

Raw scores

1. for converting to derived scores (scaled scores, percentile ranks, stanines and grade equivalents)
2. for statistical analysis such as computing coefficients of correlation between scores on two tests
3. for quick grouping of small groups of students for instructional purposes.

Grade equivalents

1. for interpreting performance of groups such as class, grade, school, district, etc.
2. for determining the relative status of individuals, especially those in the average range.

Percentile ranks

1. for determining the standing of a student on a test in relation to a specified group (local or national)
2. for comparing results among tests in a battery
3. for reporting results to parents, other laypersons, and students.

Stanines

1. same as for percentile ranks
2. for making comparisons with other variables such as academic aptitude.

Scaled scores

1. for studying change in achievement over a period of time
2. for interpreting results from out-of-level testing
3. for most statistical analyses
4. for comparing results when different battery levels and/or forms of a test have been given.

Anticipated achievement scores

1. for determining student achievement relative to potential
2. for measurement of group achievement related to group potential
3. for analysis of effectiveness of instruction relative to potential.

WHO IS RESPONSIBLE FOR TEST PROGRAM MANAGEMENT AND DATA DISSEMINATION?

The responsibilities for program management and data dissemination can be divided into three categories: Central Office, Individual School Plant and Individual Teachers. Each category has responsibility for, (1) management and (2) curriculum evaluation and application.

CENTRAL OFFICE

Management Responsibilities

1. the storage and organized distribution of tests and materials
2. maintenance of test security
3. determining district testing and reporting schedule
4. training district test coordinator and clerical staff
5. training clerical personnel at the schools for management and distribution of materials
6. training principals and/or instructional leaders at each school so that they can interpret and analyze data for instruction purposes and dissemination of data
7. training teachers to be knowledgeable about test objectives, test content, test administration and data dissemination.

Curriculum Responsibilities

1. curriculum and program analysis and evaluation
2. statistical reporting to all applicable agencies
3. predictions for program needs and improvements
4. dissemination of data analysis to schools for application
5. special program evaluation and analysis.

INDIVIDUAL SCHOOLS

Management Responsibilities

1. establish school testing schedule

2. provide organization for distribution of tests and materials

3. provide system of recording test data on student records

4. provide procedure for check and collection of tests and testing materials for return to Central Office.

Curriculum Responsibilities

1. interpret data for program evaluation and adjustment
2. provide leadership in data evaluation and dissemination
3. interpret data for grade or group progress assessment
4. evaluate instructional effectiveness
5. establish school and teacher objectives.

INDIVIDUAL TEACHERS

Management Responsibilities

1. student preparedness for testing
2. establish optimum testing schedule
3. be knowledgeable of test administrative procedures
4. administering the tests
5. collect and check all tests and testing materials for return to school office.

Curriculum Responsibilities

1. class or program reorganization based on data received
2. individual pupil evaluation for instructional purposes
3. dissemination of student data to parents. Test data should never be given to parents without personal explanation and interpretation.

The management of any testing program may be the key to its success. Teachers and school administrators should be spared as much as possible from these time-consuming tasks.

Proper orientation of all staff members regarding the purposes and procedures of the total testing program will make the entire testing program more meaningful to all individuals involved in the testing program.

REFERENCES

- Anderson, S.B., Ball, S., Murphy, R.T., & Associates.
Encyclopedia of educational evaluation. Jossey-Bass,
1977.
- Bachman, A.W., Hall, B.W., Mazur, J.L., & Ward, A.W.
Guide for school testing. National council of
measurement in education. 1977.
- California achievement tests - Technical bulletin.
CTB/McGraw Hill, 1979.
- Cox, C.H. Basic skills through mastery learning. Trends
in Education, December 1979, 363-365.
- McClung, M.D. Legal and educational issues in competency
testing. The Education Digest, January 1980, 3-5.
- Metropolitan achievement test - Teacher manual.
- Mitchell, B.C. A Glossary of measurement terms. Test
service notebook 13. Harcourt Brace, Jovanovich, Inc.
- Wargo, M.J., & Green, D.R. Achievement testing of
disadvantaged and minority students for educational
program evaluation. CTB/McGraw Hill, 1977.

THE IMPLEMENTATION OF
CRITERION-REFERENCED TESTING IN SCHOOLS

By Marc S. Becker, Assistant Director
Educational Management System
Glendale Union High School District
Glendale, Arizona

A widespread and pervasive decline in student achievement has characterized American education in the past decade. In state after state, taxpayers and their elected representatives have expressed alarm at the large proportion of students who fail to learn basic skills. Mounting public concern has led to the enactment of legislation which holds schools accountable for the improvement of student learning.

In this era of accountability, testing has assumed an importance heretofore unparalleled. The public looks upon test results as an indication of the extent to which schools are meeting their responsibility to educate students. District administrators and classroom teachers are seeking ways to use test results for improving instructional programs. State legislatures are mandating that testing be accorded priority status in school districts. Wherever one turns in the educational scene, testing is a topic of concern and lively discussion.

Educational policymakers who are charged with the implementation of accountability inevitably face an important question: What type of testing should be used to measure student achievement? A few years ago, this question would not have been asked. Only one type of testing was available--the norm-referenced type in which student achievement is compared with national norms. Within the last decade, an alternative approach to testing has emerged throughout the nation. This approach to testing compares student achievement with well-defined standards. It is known as "criterion-referenced testing."

This section of the resource guide is designed to acquaint policymakers and practitioners with criterion-referenced testing, and to provide some practical guidelines for the implementation of criterion-referenced testing in schools. The following questions will be addressed:

- . What is a criterion-referenced test?
- . How do criterion-referenced tests differ from norm-referenced tests?
- . Why use criterion-referenced tests?
- . What does a criterion-referenced testing program consist of?
- . What are the steps involved in setting up a criterion-referenced testing program?
- . What options are available for setting up a criterion-referenced testing program?

WHAT IS A CRITERION-REFERENCED TEST?

A criterion-referenced test compares student performance with standards for acceptable achievement. It is characterized by five essential attributes:

1. An area of knowledge is divided up into a set of skills.
2. Each skill is expressed as an instructional objective.
3. A standard for acceptable achievement is established for each skill.
4. The performance of an individual student in each skill is compared with the standard for acceptable achievement.
5. A student is considered competent in an area of knowledge if the student has mastered all or most of the skills which comprise it.

Criterion-referenced tests are often referred to as COMPETENCY-BASED TESTS. They are given this name because the achievement of individual skills indicates a student's competence in an area of knowledge.

EXAMPLE: Criterion-referenced Testing in Basic English

Area of knowledge:	Grammar
Component skills:	Parts of Speech Subject-verb Agreement Complements Prepositional Phrases
Skills expressed as an instructional objective:	<u>Skill: Prepositional Phrases</u> Given simple or compound sentences, each showing a group of words underlined, the student will identify the sentence in which the group of words underlined is a prepositional phrase.
Standard for acceptable achievement in each skill:	<u>Skill: Prepositional Phrases</u> A student will be considered proficient in prepositional phrases if the student has correctly answered 75% of the items which measure it.
Standard for competence in area of knowledge:	A student will be considered competent in grammar if the student has mastered all four of its component skills.

EXAMPLE: Criterion-referenced Testing in Basic Math

Area of knowledge:	Basic measurement
Component skills:	Using English measurement Reading metric instruments Reading and interpreting graphs
Skill expressed as an instructional objective:	<u>Skill: Reading and interpreting graphs</u> Given a set of data and four graphs, the student will select the graph that accurately plots the data.
Standard for acceptable achievement in each skill:	<u>Skill: Reading and interpreting graphs</u> A student will be considered proficient in reading and inter- preting graphs if the student has correctly answered 80% of the items which measure it.
Standard for competence in area of knowledge:	A student will be considered competent in basic measurement if the student has achieved all three of its component skills.

HOW DO CRITERION-REFERENCED TEST DIFFER FROM NORM-REFERENCED TESTS?

Many educators and members of the public fail to grasp the distinctions between criterion-referenced and norm-referenced testing. It is common to hear the two types of testing referred to as if they served the same purposes, or shared the same characteristics. Much confusion can be eliminated if the basic differences are understood.

<u>COMPARISON OF CRITERION-REFERENCED TEST</u> <u>FROM NORM-REFERENCED TEST</u>		
	<u>Criterion-referenced</u> <u>Test</u>	<u>Norm-referenced</u> <u>Test</u>
Purpose	To determine whether each student has achieved specific skills or concepts. To find out how much students know before instruction begins and after it is finished.	To rank each student with respect to the achievement of others in broad areas of knowledge.* To discriminate between high and low achievers.
Score Interpretation	Each individual is compared with a pre-set standard for acceptable achievement. The performance of other examinees is irrelevant. Student achievement is reported for individual skills. A student's score is usually expressed as a percentage.	Each individual is compared with other examinees and assigned a score - usually expressed as a percentile, a grade equivalent score, or a stanine. Student achievement is reported for broad skill areas.*

	<u>Criterion-referenced Test</u>	<u>Norm-referenced Test</u>
Content	Measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts. Each skill is expressed as an instructional objective.	Measures broad skill areas sampled from school district syllabi, textbooks and the judgement of curriculum experts.
Item Characteristics	Each skill is tested by at least four items in order to obtain an adequate sample of student performance and to minimize the effect of guessing. The items which test any given skill are parallel in difficulty.	Each skill is usually tested by less than four items. Items vary in difficulty. Items discriminate between high and low achievers.

The differences outlined are discussed in many texts on testing. The teacher or administrator who wishes to acquire a more technical knowledge of criterion-referenced testing and its norm-referenced counterpart, may find the following book particularly helpful.

Popham, James W. Educational evaluation. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1975.

* Some norm-referenced tests report student achievement for individual skills. These tests are described on page 52.

IMPORTANT!

A semantic confusion is widespread in American schools. Teachers, administrators and parents often use the phrase "standardized test" when referring only to norm-referenced measures. This usage has fostered the misconception that criterion-referenced tests are not standardized. Nothing could be farther from the truth. Consider the following points:

- . Standardization simply refers to uniform procedures for test administration, scoring and interpretation. It is not an attribute of a test itself.
- . A criterion-referenced test can be administered, scored and interpreted with the same degree of uniformity as its norm-referenced counterpart.

Next time someone discusses a "standardized test," ask whether she/he is referring to a norm-referenced test or to a criterion-referenced test.

WHY USE CRITERION-REFERENCED TESTS?

Teachers and administrators have long been accustomed to the use of norm-referenced tests for analyzing and interpreting student achievement. It is no exaggeration to say that a majority of school districts in Arizona and elsewhere are committed to the use of norm-referenced tests. In Arizona, this commitment is likely to continue in view of the recent accountability legislation. One may well ask why any district should consider criterion-referenced testing, particularly when norm-referenced tests are already in use. The answer to this question depends upon the ways in which teachers and administrators use test results. If they want to compare student achievement in the district with national norms, then a norm-referenced test is clearly appropriate. It is designed specifically to do this. If they want to use test results to make decisions about the placement or remediation of students, the evaluation of the local curriculum and the improvement of instruction, then a criterion-referenced test is preferred. This is true for the following reasons:

1. Criterion-referenced tests provide specific information about student achievement. A criterion-referenced test reports student achievement in specific, well-defined skills. Teachers, administrators, and parents can determine exactly what a student knows, simply by inspecting his or her performance in each skill. This interpretation is not possible in most norm-referenced tests because student achievement is reported in broad skill areas.*

2. Criterion-referenced tests facilitate decisions about the placement, promotion, and remediation of students in the local curriculum. A good criterion-referenced test accurately measures the skills or concepts which make up a local curriculum. (Test specialists refer to this type of accuracy as "content validity.") Upon examining test results, it

* Some norm-referenced tests report student achievement for individual skills. These tests are described on page 52.

is possible to determine what skills a student has or has not achieved. This information can be used to make decisions governing the following:

- . placement into the appropriate level of instruction
- . provision of remedial instruction
- . promotion from one level of instruction to the next

A norm-referenced test measures broad skill areas sampled from textbooks and school district syllabi from around the nation. It is not designed to correspond to the content of instruction in any one particular school district. Therefore, decisions about the placement, promotion and remediation of students are difficult to make when using norm-referenced tests alone.

3. Criterion-referenced tests are especially well-suited for the evaluation of instruction in the local curriculum. Criterion-referenced test results show the proportion of students achieving the skills within the local curriculum. This information helps teachers and administrators determine whether instruction has been effective or needs improvement. It enables policymakers to decide if a curriculum program should be continued, revised or terminated.

A norm referenced test is difficult to use for the evaluation of local curriculum programs for the following reasons:

- . The content of a norm-referenced test differs to a small or large extent from the content of the local curriculum.
- . Student achievement in a norm-referenced test is compared with national norms.
- . National norms fluctuate. In the past decade, they have shifted downward.
- . The fact that a majority of students within a district score above national norms does not indicate high quality instruction if the norms are relatively low.

4. Criterion-referenced test results can be used to improve instruction. The skills which are not achieved by a large proportion of students may point to areas of the curriculum where instruction needs to be improved. Upon examining test results, teachers are able to turn their attention to making the necessary improvement.

An illustration of the types of decisions that can be made about students and the local curriculum when using criterion-referenced tests appears on the following page.

DECISIONS BASED UPON CRITERION-REFERENCED TESTING
MATHEMATICS PROGRAM - CACTUS COVE HIGH SCHOOL DISTRICT

Decision Context	Decisions Made
1. Placement into the appropriate level of instruction	1a. Entering freshmen were given a criterion-referenced pre-test during the first week of school. Students were tested in 16 basic arithmetic skills in six areas: whole numbers, fractions, decimals, percents, measurement and graphs. 1b. Entering freshmen not proficient in 10 of the 16 basic arithmetic skills were placed in Basic Math. 1c. Entering freshmen who were proficient in more than 10 of the 16 basic arithmetic skills were subsequently given a criterion-reference pretest in algebra. The results of this test were used to place each student in one of three courses: Basic Algebra, Algebra 1-2, or Algebra 3-4.
2. Provision of remedial instruction	2. Upon inspecting pretest results, a teacher discovered that 6 students in her Basic Algebra class were not proficient in the division of fractions. Before introducing a unit which requires the use of fractions, she gave these students remedial workbooks and practice exercises.
3. Promotion of students from one level of instruction to the next	3. A criterion-referenced posttest was given at the end of the school year to students enrolled in each math course. Test results (along with teacher recommendation) were used to determine if students should be promoted from each course into the next level of instruction (e.g., students who mastered 12 of the 14 skills in Algebra 1-2 were promoted to Algebra 3-4).

Decision Context	Decisions Made
4. Continuation, termination, or revisions of instructional programs	4. Upon inspecting posttest results, administrators and board members found that 80% of the students enrolled in the Basic Math achieved all skills. They recommended that Basic Math continue to be implemented without modification.
5. Improvement of instruction	5. Upon inspecting posttest results, teachers of Basic Math learned that 42% of their students failed to master the computation of percents. They worked together in developing a plan for improving instruction in the computation of percents for next year.

The variety of decisions which can be based upon the use of criterion-referenced tests make them a powerful tool for implementing accountability. A district which is committed to accountability would do well to supplement norm-referenced testing with a criterion-referenced testing program. In the section which follows, the components of a criterion-referenced testing program are described.

WHAT DOES A CRITERION-REFERENCED TESTING PROGRAM CONSIST OF?

A criterion-referenced testing program should contain, at minimum, the following components:

- . pre-and posttesting
- . uniform procedures for test administration
- . item security
- . student achievement reports

1. Pre-and posttesting. In order to be useful to teachers and administrators, a criterion-referenced test should facilitate decisions about the individual students, and about the curriculum. These decisions can best be made if testing occurs in the fall before instruction begins, and in the spring after it is finished. Pre-and posttesting is clearly desirable.

a. Pretest results in the fall:

- . reveal what students already know and what they don't know.
- . help teachers plan instruction to meet the needs of individual students and groups of students.
- . enable administrators to work with teachers in placing students at the level of instruction appropriate to their skills and capabilities.

b. Posttest results tell teachers and administrators:

- . how much students have learned as the result of instruction.
- . whether instruction has been successful.
- . which students are ready for promotion to the next level of instruction.
- . which students need to be retained for remediation.

2. Uniform procedures for test administration. The reliability of a test can be seriously impaired if the procedures for test administration differ from setting to setting within a district. If more than one setting is utilized, it is imperative that the settings for test administration be similar, that time constraints be uniform, and that the directions provided to students be identical.

3. Item security. It is good practice to guard against the dissemination of test items to anyone not connected with test development. When items become available to the public, they ultimately end up in the hands of students prior to testing.

4. Student achievement reports. If criterion-referenced testing is to be useful, test results must be made available to students, teachers, parents and administrators. Reports may indicate student achievement within classrooms, schools, programs, and across the district as a whole. A district may wish to purchase these reports or to design its own.

WHAT ARE THE STEPS INVOLVED IN SETTING UP
A CRITERION-REFERENCED TESTING PROGRAM?

The establishment of a criterion-referenced testing program requires the following steps:

- . obtain local commitment
- . establishing a planning committee
- . deciding upon the curriculum which is to be the focus of testing
- . developing a structure for each curriculum program in which testing occurs
- . deciding upon the skills to be tested
- . developing or purchasing the test instrument.

Obtaining local commitment. Step 1: The best-intentioned effort to initiate criterion-referenced testing will fail unless the commitment of policymakers and teachers is obtained. This commitment can only be obtained if administrators and teachers understand the basic purposes and potential uses of criterion-referenced tests and how they differ from norm-referenced tests. Inservice activities designed to promote understanding of criterion-referenced testing is one way of obtaining commitment.

Establishing a planning committee. Step 2: In order to be of practical use to a school district, a criterion-referenced test must measure the skills or concepts taught in the local curriculum. It is therefore necessary for teachers and curriculum staff to participate in decisions governing the content and focus of testing (steps 3-6). Teacher participation may be secured through the establishment of a planning committee which also includes principals and district administrative staff. It is important that the teachers serving on the committee be selected in such a way that each school within the district is represented.

Deciding upon the curriculum which is to be the focus of testing. Step 3: The establishment and implementation of a testing program can be a costly affair for any school district. A district with limited financial resources may want to initiate testing in one curriculum program only. If additional funds become available, the district then may want to consider the expansion of testing to other programs. The program initially selected for testing should be that in which student achievement appears to be the lowest.

Developing a structure for each curriculum program in which testing occurs. Step 4: A program structure consists of the

skills and concepts which make up a designated local curriculum. The development of a program structure requires the extensive participation of teachers. It involves the following steps:

- a. the formulation of instructional goals
- b. identification of exit skills, skills or concepts which students are expected to learn by the time they have finished instruction
- c. identification of enabling skills, prerequisite skills or concepts which "enable" students to achieve exit skills
- d. identification of entry level skills, skills or concepts which students should have learned before entering a course or before beginning a new unit of instruction
- e. establishing a sequence in which the skills should be learned and taught
- f. expressing each skill as an instructional objective.

The development of a program structure has the following advantages for testing and instructional planning:

- . the skills outlined in the program structure help teachers and administrators make decisions about the content of tests
- . teachers are able to use the program structure as a vehicle for planning instruction which will help students achieve the skills tested
- . teachers can use the program structure to make improvements in instruction for skills in which students are weak.

One of the most effective procedures for developing a program structure is the use of an instructional task analysis. An instructional task analysis is a process by which the sequence of skills is analyzed and depicted visually. Practitioners who are interested in becoming acquainted with instructional task analysis may find the following reference helpful:

Dick, Walter, & Carey, Lou. The systematic design of instruction. Glenview, Ill, Scott Foresman and Co., 1978, 24-76.

Deciding upon the skills to be tested. Step 5: Upon examining the program structure, teachers and administrators can decide which skills should be the focus of pre-and posttesting.

Developing or purchasing the test instrument. Step 6: A central task of the planning committee is to analyze the resources and constraints which exist within the district. As the result of this analysis, the committee will be able to recommend one of the following options for the purchase or development of the test instrument:

- . the purchase of ready-made criterion-referenced tests
- . test customization
- . the purchase of instructional objectives for test development
- . local development of criterion-referenced tests.

Each of these options is described in the following pages.

OPTION: THE PURCHASE OF A READY-MADE CRITERION-REFERENCED TEST

DESCRIPTION

A number of publishing companies are currently marketing criterion-referenced tests developed by their own staff of specialists. The tests measure skills in a variety of subject areas and grade levels. In physical appearance, they resemble their norm-referenced counterparts in that students are provided with test booklets and answer sheets. A uniform set of directions is provided for test administrators.

ADVANTAGES

- . Objectives and test items are written by the publisher.
- . No special expertise is required to implement testing.
- . Costs are within the range of most school district budgets.

DISADVANTAGES

- . Teachers and administrators must agree that the skills measured on the test are important.
- . The skills measured on the test may not correspond completely to the skills which make up the local curriculum.
- . If the skills in the test differ from the skills which make up the local curriculum, the test will not be practical for the placement or remediation of students, the evaluation of the curriculum program or the improvement of instruction.
- . The test may have been developed with students who differ from local students in ethnic background and socio-economic status.

MATERIALS TO BE PURCHASED

- . test booklets
- . answer sheets: hand scored or machine scored.

OPTIONAL SERVICES*

- . computer scoring of answer sheets
- . preparation of test reports for students, teachers, parents and administrators
- . reporting of student achievement by individual skill, classroom, school, program and district
- . resource guides for teachers that match textbook references to each skill tested
- . instructional materials which teachers may use to remediate students in each skill tested
- . classroom unit tests which may be used between district pre-and posttesting.

HOW TO SHOP FOR A CRITERION-REFERENCED TEST

- . Obtain sample test materials from more than one publishing company. Sample test materials are usually available without charge.
- . Examine sample test materials to determine which test may be most suited to the local curriculum in terms of grade level and content. The sample materials should include a list of the skills tested and their corresponding objectives.
- . Contact the regional representative of the company that published a test of interest to the district.
- . Obtain a technical manual which explains how the test was developed.

WHAT TO LOOK FOR IN A CRITERION-REFERENCED TEST

In order to be useful to a district, a criterion-referenced test must meet all of the following requirements:

- . Skills and concepts are expressed as instructional objectives.
- . The skills tested correspond to the skills which make up the local curriculum.
- . Each skill is tested by at least four items.
- . The items which test any given skill are parallel in difficulty.

* The optional services available to test purchasers vary from company to company.

- . The test was developed with students who were similar to local students in socio-economic status and ethnic composition.

The best way to find out if a test meets all of these requirements is to talk to the regional representative of the company which publishes it.

IMPORTANT!

It is possible to purchase norm-referenced tests which offer a criterion-referenced interpretation of student achievement. These tests report student performance for individual skills and may specify standards for acceptable achievement. It is important to know that these tests are not criterion-referenced for two reasons:

- . The number of items testing each skill is usually less than four.
- . Items testing each skill vary in difficulty and discriminate between high and low achievers.

A norm-referenced test which offers a criterion-referenced interpretation may be useful for comparing student achievement with national norms and for general instructional planning. It does not provide the precise information necessary to make decisions about the placement and promotion of students, the evaluation of the local curriculum or the improvement of instruction.

WHOM TO CONTACT

Below is a partial list of major publishing companies which market criterion-referenced tests.

CTB McGraw Hill Inc.
1700 E. Don Carlos #1-U
Tempe, AZ 85281
Phone 602 894-1926

Science Research Associates
155 N. Wacker Drive
Chicago, IL 60606

Instructional Objectives
Exchange (IOX)
Box 24095 Dept T
Los Angeles, CA 90024
Phone 213 988-1548

Scott, Foresman and Company
11728 Darlington Ave. #8
Los Angeles, CA 90049
Phone 213 826-8574

The Psychological Corporation
757 Third Avenue
New York, NY 10017

OPTION: CRITERION-REFERENCED TEST CUSTOMIZATION

DESCRIPTION

Within recent years, a new alternative has become available to school districts wishing to implement criterion-referenced testing. A growing number of test publishers are willing to send their own specialists to work in school districts as consultants. These consultants collaborate with teachers and administrators in the development of criterion-referenced testing programs suited to the characteristics of the local student population and to the curriculum. This collaborative effort, known as "test customization," can take place at any grade level.

ADVANTAGES

- . A customized test directly measures the skills and concepts which make up the local curriculum.
- . Teachers and administrators are extensively involved in the test development effort. As a consequence, they are likely to experience "ownership" of the testing program.
- . Due to their involvement in test development, teachers will be able to plan instruction more effectively.
- . The presence of skilled testing specialists insures that the test(s) developed are high in quality.
- . The tests are developed using the local student population.
- . Given their match to the curriculum, customized tests are highly practical for the placement or remediation of students, the evaluation of the curriculum program and for the improvement of instruction.

DISADVANTAGES

- . The process of test customization can be time consuming.
- . Test customization can be costly, depending upon the services which are contracted.

NOTE

It is difficult to state any generalizations about the costs of customization since these vary from company to company. A district considering test customization should contact more than one company to obtain complete information about costs and services provided.

THE CUSTOMIZATION PROCESS

One or more consultants collaborate with district personnel in:

- . development of a curriculum structure containing goal statements and a sequence of skills
- . expression of skills as instructional objectives
- . selection of skills to be tested
- . setting standards for acceptable student achievement in each skill tested
- . selection of four or more items from an existing item bank to test each skill
- . modification of items for adaptation to the local curriculum and/or the cultural composition of the student population
- . preparation of test booklets for use by students.

The first three activities listed above may be completed by district personnel prior to contracting for test customization.

OPTIONAL SERVICES

- . Provision of machine scored and hand scored answer sheets.

- . Computer scoring of answer sheets.
- . Preparation of test reports for students, teachers, administrators and parents.
- . Reporting of student achievement for individual students, classrooms, schools and for the district.

WHOM TO CONTACT

Below is a partial list of companies which offer test customization services.

CTB McGraw Hill Inc.
1700 E. Don Carlos #1-U
Tempe, AZ 85281
Phone 602 894-1926

EPIC Instructional Management
System, Inc.
P.O. Box 18676
Tucson, AZ 85731
Phone 602 886-0351

Educational Testing
Service
Princeton, NJ 08540

Instructional Objectives
Exchange (IOX)
Box 24095 Dept T
Los Angeles, CA 90024

OPTION: THE PURCHASE OF INSTRUCTIONAL OBJECTIVES
FOR CRITERION-REFERENCED TEST DEVELOPMENT

DESCRIPTION

Collections of instructional objectives are available for purchase from some test publishers. The objectives correspond to skills and concepts in various subject areas and at several grade levels. Each collection of objectives is usually packaged in a booklet. A district may purchase a booklet and use the objectives that it contains to plan a curriculum or to develop criterion-referenced tests.

ADVANTAGES

- . The objectives correspond to skills and concepts in various subject areas and at several grade levels.
- . A district is given a wide range of objectives from which to choose.
- . A task of writing instructional objectives is assumed by the publisher, not by district personnel.
- . The objectives provide guidance for writing test items (objectives may be accompanied by one or more sample items).
- . The objectives may be used as is or modified to be more congruent with the local curriculum.
- . The purchase of instructional objectives is relatively inexpensive.

DISADVANTAGES

- . Teachers and curriculum staff must assume all of the tasks involved in test development (described on page 63).
- . A testing specialist must be available or accessible to guide and supervise all test development activities.
- . The process of test development requires a considerable investment of time.
- . Teachers who participate in test development may have to be released from the classroom for periods of time.

- . The time and personnel required for test development may prove costly.

MATERIALS TO BE PURCHASED

Each collection of instructional objectives is referenced to a particular grade level and subject area (e.g., 5th grade reading). Each collection of objectives is purchased in the form of a booklet. A catalog of instructional objective collections may be obtained without charge from the publishing company.

WHAT TO LOOK FOR

It may not be possible to inspect the objectives until they have been purchased from the publisher. Upon receiving the objectives, it is necessary to decide whether they can be used to develop good criterion-referenced tests. In order to be usable, each objective must:

- . describe a specific, testable behavior
- . correspond to a skill or concept in the local curriculum.

If the purchased objectives do not meet either or both of these requirements they should be modified or not used.

WHOM TO CONTACT

Commercial-Educational
Distributing Services
P.O. Box 8723
Portland, OR 97208

Educational Testing Service
Princeton, NJ 08540

Institute for Educational
Research
1400 W. Maple Ave
Downers Grove, IL 60515

Instructional Objectives
Exchange (IOX)
Box 24095 Dept T
Los Angeles, CA 90024

Houghton-Mifflin Company
Test Dept.
777 California Ave
Palo Alto, CA 94304

Science Research Associates
259 E. Erie St.
Chicago, IL 60611

OPTION: LOCAL DEVELOPMENT OF CRITERION-REFERENCED TESTS

DESCRIPTION

All components of test development are completed by teachers, curriculum staff and administrators within a district. A testing specialist guides and supervises all test development activities.

ADVANTAGES

- . A locally developed test directly measures the skills and concepts which make up the local curriculum.
- . Teachers and administrators are extensively involved in the test development effort. As a consequence, they are likely to experience "ownership" of the testing program.
- . Due to their involvement in test development, teachers will be able to plan instruction more effectively.
- . The presence of a skilled testing specialist insures that the tests developed are high in quality.
- . The tests are developed using the local student population.
- . Given their match to the curriculum, locally developed tests are highly practical for the placement or remediation of students, the evaluation of the curriculum program and for the improvement of instruction.

DISADVANTAGES

- . Local development of criterion-referenced tests requires a considerable expenditure of time and human resources.
- . The test development process is costly.
- . Teachers and curriculum staff must assume all of the tasks of test development.
- . Teachers who participate in test development may have to be released from the classroom for periods of time.

- . A testing specialist must be available or accessible to guide and supervise all test development activities.
- . Test booklets must be assembled and printed by the district.
- . A computer facility for scoring tests should be available or accessible.
- . Student achievement reports must be developed by the district.

THE TEST DEVELOPMENT PROCESS

The test development process consists of the following activities:

- . development of a curriculum structure containing goal statements and a sequence of skills
- . expression of skills as instructional objectives
- . selection of skills to be tested
- . setting standards for acceptable student achievement in each skill tested
- . writing a bank of items for each skill tested
- . fieldtesting of items to increase test reliability
- . selection of four or more items from the item bank to test each skill
- . preparation of test booklets for use by students
- . design or purchase of answer sheets
- . scoring of answer sheets
- . development of student achievement reports.

WHOM TO CONTACT

The Glendale Union High School District has been developing criterion-referenced tests for local use since 1972. Information about the test development process can be obtained by contacting:

Dr. Gerald E. George, Director or
 Mr. Marc S. Becker, Assistant Director
 Glendale Educational Management System
 Glendale Union High School District
 7650 N. 43rd Avenue Glendale, AZ 85301

REFERENCES

Dick, Walter, & Carey, Lou. The Systematic design of instruction. Glenview, Illinois: Scott Foresman and Co., 1978.

Popham, James W. Educational evaluation. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1975.

APPENDIX A .

GLOSSARY

GLOSSARY

Academic Aptitude - An estimate of probability of school success based on measures of acquired and native abilities. (Also academic potential, scholastic aptitude.)

Battery - A group of tests, usually standardized on the same sample population so that test results are comparable.

Content Validity - The extent to which a test adequately measures the skills and concepts which make up a designated curriculum.

Criterion-referenced Test - A test designed to measure student achievement in the skills and concepts which make up a designated curriculum. The performance of each student is compared to a standard for acceptable achievement in each skill tested. The performance of other examinees is not considered in score interpretation.

Deviation - The amount by which a score differs from some reference value, such as the mean, the norm, or the score on some other test.

Diagnostic Test - A test usually in specific skill areas used to analyze and locate an individual's specific areas of strength or weakness and, in some instances, determine their cause.

Empirical Norming Date - The specific date upon which a test was administered to the sample population to establish the norms.

Fieldtesting - The process whereby newly written test items are tried out on students and revised.

Goal Statement - A broad description of the skills or concepts which students are expected to learn as the result of instruction.

Instructional Objective - A specifically described skill or concept which students are expected to learn as the result of instruction. An instructional objective identifies an expected student behavior and the conditions under which the behavior is expected to occur.

Instructional Validity - The concept that the learning objectives that are tested have been taught.

Item - A single question or exercise in a test.

Item Analysis - Evaluating single test items to determine relationship and often its correlation with some external criterion such as a learning objective.

Item Bank - A group of test items written for each skill tested. The items are parallel in format and difficulty. On any occasion for testing, a subset of items is sampled from the bank and put into the student test booklet.

Percentile Band - The interpretation of a test score that takes into account the measurement error. The range of such bands is usually from one standard error of measurement below the obtained score to one above it.

Test Reliability - The extent to which a test consistently measures what it is intended to measure.

APPENDIX B

HOUSE BILL 2013

ARTICLE 2.1.

READING, GRAMMAR AND MATHEMATICS ACHIEVEMENT TESTS

15-1131. Testing of pupils

A. THE STATE BOARD OF EDUCATION SHALL:

1. PROMULGATE RULES AND REGULATIONS PURSUANT TO THIS ARTICLE.
2. ADOPT AND IMPLEMENT BY JANUARY 1, 1981 A STATEWIDE
NATIONALLY STANDARDIZED NORM-REFERENCED ACHIEVEMENT TESTING PRO-
GRAM FOR USE IN GRADES ONE THROUGH TWELVE.

3. ADOPT A NATIONALLY STANDARDIZED NORM-REFERENCED ACHIEVE-
MENT TEST WHICH SHALL BE UNIFORM THROUGHOUT THE STATE.

4. INSURE THAT THE RESULTS OF TESTS ESTABLISHED AS PROVIDED
IN THIS ARTICLE ARE COMPARATIVE TO ASSOCIATED GRADE EQUIVALENTS,
PERCENTILES AND STANINES DERIVED FROM A MULTISTATE SAMPLE.

B. THE TESTING PROGRAM SHALL BE DESIGNED TO TEST PUPILS IN
GRADES ONE THROUGH TWELVE IN READING, GRAMMAR AND MATHEMATICS AT
A LEVEL APPROPRIATE FOR THEIR GRADE LEVEL.

C. THE STANDARDIZED NORM-REFERENCED ACHIEVEMENT TESTS ADOPTED
BY THE BOARD SHALL BE GIVEN ANNUALLY IN A WEEK IN APRIL AS DETER-
MINED BY THE STATE BOARD.

D. A TEST FOR PENMANSHIP SHALL NOT BE REQUIRED PURSUANT TO
THIS ARTICLE.

15-1132. Acceptable tests; contract for purchase of tests

A. THE STATE BOARD OF EDUCATION SHALL ENTER INTO CONTRACTS
WITH CONTRACTORS FOR THE PURCHASE OF THE TESTS ADOPTED BY THE STATE
BOARD. THE CONTRACTS MAY ALSO PROVIDE FOR THE DISTRIBUTION OF THE
TESTS TO THE SCHOOL DISTRICTS AND THE SCORING OF THE TESTS.

CONTRACTORS SHALL GIVE A CASH OR CORPORATE SURETY BOND PAYABLE TO
THIS STATE AND APPROVED BY THE BOARD INDEMNIFYING THE STATE IN THE
TEST PURCHASES IN AN AMOUNT NOT LESS THAN FIVE HUNDRED NOR MORE
THAN TEN THOUSAND DOLLARS AS MAY BE DETERMINED BY THE BOARD IF:

1. THE CONTRACTOR WILL FAITHFULLY COMPLY WITH THE CONDITIONS
OF THE CONTRACT AND WILL FURNISH TO THE STATE THE TESTS AS PROVIDED
IN THE CONTRACT AT PRICES NOT EXCEEDING THE LOWEST PRICES THEN
GRANTED TO ANY BUYER.

2. THERE IS A DECREASE IN THE PRICES GIVEN TO A PERSON
PURCHASING SUCH TESTS FROM THE CONTRACTOR, THE STATE SHALL HAVE
THE BENEFIT OF THE DECREASE IN PRICE.

B. THE CONTRACTOR SHALL FILE WITH THE BOARD A SWORN STATEMENT
STATING THE LOWEST PRICES FOR WHICH THE CONTRACTOR'S SERIES OF
TESTS IS SOLD ANYWHERE IN THE UNITED STATES.

C. IF A CONTRACTOR VIOLATES A CONDITION OF THE CONTRACT, THE
ATTORNEY GENERAL SHALL, UPON REQUEST OF THE STATE BOARD OF EDUCA-
TION, INSTITUTE AN ACTION FOR DAMAGES ON THE BOND OF THE CONTRACTOR.

15-1133. Test results

A. WHEN THE CONTRACTOR PROVIDES FOR THE SCORING OF THE TESTS
HE SHALL REPORT SUMMARY RESULTS OF TESTS ADMINISTERED TO PUPILS AS
PROVIDED IN THIS ARTICLE TO THE STATE BOARD OF EDUCATION AND THE
COUNTY SCHOOL SUPERINTENDENTS AND THE STATE BOARD SHALL MAKE THE
RESULTS AVAILABLE TO THE PUBLIC. WHEN THE STATE BOARD OF EDUCATION
PROVIDES FOR THE SCORING OF THE TESTS THE STATE BOARD SHALL REPORT

SUMMARY RESULTS OF TESTS ADMINISTERED TO PUPILS AS PROVIDED IN THIS ARTICLE TO THE COUNTY SCHOOL SUPERINTENDENTS AND SHALL MAKE THE RESULTS AVAILABLE TO THE PUBLIC. THE SUMMARY RESULTS SHALL INCLUDE FOR EACH INDIVIDUAL PUPIL, SCHOOL, SCHOOL DISTRICT, COUNTY AND THE STATE:

1. THE TEST SCORES BY SUBJECT AREA.
2. THE NUMBER OF PUPILS WHO TOOK THE TEST BY GRADE LEVEL AND SUBJECT AREA AND THEIR AVERAGE READING, GRAMMAR AND MATHEMATICS SCORES.
3. SUCH OTHER INFORMATION OR COMPARATIVE DATA AS THE STATE BOARD OF EDUCATION MAY, BY REGULATION, REQUIRE.
- B. TEST RESULTS ON INDIVIDUAL PUPILS SHALL NOT BE MADE AVAILABLE TO THE PUBLIC BY NAME OR INDIVIDUALLY IDENTIFIABLE REFERENCE.
- C. THE STATE BOARD SHALL PROVIDE A COPY OF THE RESULTS FROM EACH SCHOOL DISTRICT TO THAT DISTRICT. NO RESULTS MAY BE RELEASED TO THE PUBLIC UNTIL TEN DAYS AFTER THE REPORTS ARE PROVIDED TO EACH DISTRICT.
- D. THE STATE BOARD SHALL PROVIDE EACH SCHOOL DISTRICT PARTICIPATING IN THE TESTING PROGRAM WITH A COPY OF EACH PUPIL'S SCORE IN READING, GRAMMAR AND MATHEMATICS AND THE ASSOCIATED GRADE EQUIVALENTS, PERCENTILES AND STANINES FOR THE SCHOOL, SCHOOL DISTRICT, COUNTY, STATE AND NATION.
- E. THE DISTRICT SHALL PROVIDE A PARENT OR GUARDIAN OF EACH PUPIL PARTICIPATING IN THE TESTING PROGRAM WITH A COPY OF THE PUPIL'S SCORE IN READING, GRAMMAR AND MATHEMATICS AND THE ASSOCIATED GRADE EQUIVALENTS, PERCENTILES AND STANINES FOR THE SCHOOL, SCHOOL DISTRICT, COUNTY, STATE AND NATION.
- F. THE STATE BOARD MAY ANNUALLY REVIEW THE TESTS AND MAKE APPROPRIATE CHANGES, EXCEPT THAT THE BOARD SHALL NOT CHANGE THE TESTS USED UNLESS IT IS POSSIBLE TO ACCURATELY EQUATE THE SCORES FROM THE FORMER TO THE NEW TESTS.
- G. THE SUPERINTENDENT OF PUBLIC INSTRUCTION SHALL MAINTAIN A COMPUTERIZED FILE OF TEST RESULTS FOR RESEARCH AND ANALYSIS.

15-1134. Exemption

PUPILS WHO ARE TRAINABLE MENTALLY HANDICAPPED, EDUCABLE MENTALLY HANDICAPPED, VISUALLY HANDICAPPED, HEARING HANDICAPPED, MULTIPLE HANDICAPPED, SERIOUSLY EMOTIONALLY HANDICAPPED OR WHO ARE NOT REQUIRED TO ATTEND REGULAR CLASSES IN A PUBLIC SCHOOL AS PROVIDED IN THIS TITLE ARE EXEMPT FROM THE TESTING REQUIREMENT PRESCRIBED BY THIS ARTICLE.

15-1135. Testing program costs; inclusion in budget

BEGINNING WITH THE 1980-81 FISCAL YEAR THE STATE BOARD OF EDUCATION SHALL INCLUDE WITHIN ITS BUDGET ALL COSTS PERTAINING TO THE ACHIEVEMENT TESTS PRESCRIBED IN THIS ARTICLE.

APPENDIX C

BIBLIOGRAPHY OF AEIS DOCUMENTS

BIBLIOGRAPHY OF AEIS DOCUMENTS

Journal Articles Pertaining to Criterion-Referenced Testing

EJ 188 634

"A Model for a Criterion-Referenced System." Dorothy Bray, Darrel D. Taylor. NASSP Bulletin, October 1978, p 70-78. Guides school administrators through a process to institute a workable criterion-referenced instructional system.

EJ 097 680

"Performance Objectives and C-R Tests--We Wrote Our Own." Mildred Royal. Reading Teacher, April 1974, p 701-03.

Microfiche Documents Pertaining to Criterion-Referenced Testing

ED 170 338 (2 mf)*

New Jersey Statewide Minimum Standards: Results from the Program's Second Year. Stephen L. Köffler. Occasional Papers in Education. New Jersey State Dept. of Education, Trenton. Division of Operations, Research, and Evaluation.

ED 167 503 (1 mf)*

Competency Testing. An Annotated Bibliography. Michael Jackson, Barbara Battiste. Competency testing for either graduation from high school, or as a method for assessing whether a student should advance to a higher grade level, is the focus of this annotated bibliography. Documents relating to accountability, competency testing, general testing concepts, and testing requirements are cited.

ED 117 145 (1 mf)*

Large-Scale Objective Referenced Testing: Some Practical Problems and Concerns. William J. Mathis. The primary purpose of this paper presented at a 1975 Meeting of the National Council on Measurement in Education is to examine workable methods for conducting large scale state testing or assessment programs with criterion or performance referenced measures. Methods of resolving the

the problems posed by these programs are addressed with professional and community involvement and test validity strongly emphasized.

ED 120 213 (1 mf)*

Operating an Objective-Referenced Testing Program: Florida's Approach to Large-Scale Assessment. Colorado State Dept. of Education, Denver. Cooperative Accountability Project. Sponsoring Agency: Bureau of Elementary and Secondary Education (DHEW/OE), Washington, D.C. Div. of State Agency Cooperation. This monograph provides information on how to implement objective-referenced testing, particularly the large-scale testing required for statewide testing programs. The major topics included are accountability and product objectives, contracting, training, administration problems, and analysis of results. The text describes procedures used in the 1974-75 Florida Statewide Assessment Program and offers practical suggestions throughout.

ED 116 263 (1 mf)*

Accountability and Testing. NAESP School Leadership Digest Second Series, Number 5. ERIC/CEM Research Analysis Series, Number 20. Jo Ann Mazzaella. National Association of Elementary School Principals, Washington, D.C.; Oregon University, Eugene. ERIC Clearinghouse on Educational Management. This paper assists administrators faced with the decision of what methods of evaluation to use for accountability by discussing the pros and cons of standardized, or norm-referenced, tests and of criterion-referenced tests, and by examining alternatives for assessing student achievement.

ED 162 417 (2 mf)*

The Competency Challenge: What Schools are Doing. Peggy Gorder. National School Public Relations Association, Arlington, Va. This report probes current methods of constructing, administering, evaluating, and explaining competency tests; it gives detailed accounts of specific state and local programs; it presents case histories that can enable the administrator and educator to benefit from the experience of others; and it shows samples of actual competency tests currently in use. It goes on to examine legal problems that have arisen out of competency testing, such as cultural bias and insufficient preparation, and then examines implications and trends for the future. The appendix contains National Assessment of Educational Progress sample questions, a table summarizing state activity in minimal competency testing, and other related resources.

ED 170 383 (1 mf)*

Development of a Comprehensive Proficiency Testing System.

Michael D. Hiscox. Northwest Regional Educational Lab., Portland, Oregon. Paper presented at the 1979 Meeting of the American Educational Research Association. Components of the Proficiency Testing System (PTS) designed for the Central Valley School District, Washington, are described. Components are: (1) progress assessment, consisting of norm referenced tests, writing samples, diagnostic tests, and ability checklists; (2) user assistance, including an objectives manual, inservice workshop, and a recordkeeping system; (3) essential skills consisting of criterion-referenced tests and a remedial instruction program; (4) area competencies, comprised of achievement tests in areas such as science and health, and short remedial courses; (5) survival skills, including instruments for measuring a student ability to function in modern society; and (6) gifted and talented identification procedures. A checklist of potential problems in implementing PST is also supplied.

Journal Articles Pertaining to
Norm-Referenced Testing

EJ 193 566

"Who's Asking the Questions?" Timothy Spofford. American Educator: The Professional Journal of the American Federation of Teachers, Win 1978, p 8-11. This article reviews the development of, and controversy behind, minimum competency testing at the state and national levels.

Microfiche Documents Pertaining
to Norm-Referenced Testing

ED 173 424 (1 mf)*

Handbook for the Release of Test Data. The Office of Research Report Series. Lorin W. Anderson. South Carolina State Dept. of Education, Columbia. Office of Research. School districts are presented guidelines in this report for interpreting the results, content, and purpose of testing programs that compare local students to national sample groups. Suggestions include interpreting test results by displaying percentile ranks, standard error of measurement, and frequency distributions in graphic form. Methods for releasing test data to parents and the news media are

also discussed.

ED 172 230 (1 mf)*

Assault with a Deadly Mandate: The Minimum Competency Movement.

Allan E. Dittmer. Paper presented at the combined 1979 Meeting of the Conference on English Education and the Secondary School English Conference. The author of this paper maintains that minimal competency testing is regressive and dull, distracting students from confronting important issues. Special problems surrounding this type of testing are discussed, including the "grammar" of tests as a medium, the heavy reliance of tests on math and statistics, and inaccuracy of tests as means of predicting success.

ED 170 383 (1 mf)*

Development of a Comprehensive Proficiency Testing System.

Michael D. Hiscox. Northwest Regional Educational Lab., Portland, Oregon. Paper presented at the 1979 Meeting of the American Educational Research Association. Components of the Proficiency Testing System (PTS) designed for the Central Valley School District, Washington, are described. Components are: (1) progress assessment, consisting of norm referenced tests, writing samples, diagnostic tests, and ability checklists; (2) user assistance, including an objectives manual, inservice workshop, and a recordkeeping system; (3) essential skills, consisting of criterion referenced tests and a remedial instruction program; (4) area competencies, comprised of achievement tests in areas such as science and health, and short remedial courses; (5) survival skills, including instruments for measuring a student's ability to function in modern society; and (6) gifted and talented identification procedures. A checklist of potential problems in implementing PST is included.

ED 170 285 (1 mf)*

Learning to Manage a Minimum Competency Testing Program. Robert

N. Gourley. Paper prepared for the 1979 Meeting of the American Association for Colleges of Teacher Education. This guide outlines a model for competency-based education (CBE) as well as procedures for administration of CBE programs. Topics covered include: (1) interpreting competency statements; (2) organizing instruction; (3) designing/selecting assessment instruments techniques; (4) administering the instrument/technique; (5) grading/scoring the instruments; and (6) successful and unsuccessful students.

ED 167 578 (1 mf)*

Profiles of School District Performance, 1977-78: A Guide to Interpretation. California State Dept. of Education, Sacramento. Office of Program Evaluation and Research. This interpretive guide explains the California Statewide assessment program. A brief explanation of the tests' development, content, administration, and limitations is given and a sample profile from a fictitious school district is included. Mean scores, reporting the average percent of questions answered correctly in the district for reading in grades 2, 3, 6, and 12, as well as parallel report of scores for the past 3 years are presented.

ED 164 644 (1 mf) *

Competency Testing and Handicapped Students. Merle Steven McClung, Diana Pullin. Harvard University, Cambridge, Mass. Center for Law and Education. This article discusses the fairness and legality of competency testing programs for the handicapped. Concerns addressed include: (1) exemptions for handicapped students; (2) individual determinations; (3) differential diplomas and standards; and (4) differential assessment procedures. Policymakers deciding to apply competency testing programs to handicapped students are advised to avoid any uniform approach and consider instead an approach based upon individual determinations.

ED 164 643 (1 mf) *

Competency Testing: Potential for Discrimination. Merle Steven McClung. Harvard University, Cambridge, Mass. Center for Law and Education. This article discusses the potential for discrimination inherent in competency based testing examining such topics as: (1) racial discrimination; (2) inadequate phase-in periods; (3) non-reliable and invalid tests; (4) inadequate matching of the instructional program to the test; (5) inadequate remedial instruction; and (6) unfair apportionment of responsibility between students and educators. Numerous studies and legislative results are used to illustrate this paper's argument that more valid competency testing is needed.

ED 157 947 (1 mf) *

The Costs of Legislated Minimal Competency Requirements. Prepared for the Minimal Competency Workshop sponsored by the Education Commission of the States and the National Institute of Education. Barry D. Anderson. The costs to

school districts of setting up and implementing legislated minimal competency testing (MCT) is examined. The author suggests that financial outlays are probable for items such as test development, administration, state monitoring, and remediation programs. To make testing programs cost-effective it is recommended that (1) the state rather than local agencies should administer the programs; (2) testing should not focus on the individual student; (3) cooperatives should be established to share tests; (4) test results should be reported publicly; (5) school districts should be rewarded for high performance; (6) financial inequities should be reduced; and (7) federally run testing should be discouraged. Descriptions of MCT programs by state are included.

ED 148 832 (3 mf)*

Staff Development Through Inservice Workshops in Testing for Teachers, Counselors and Principals. Kurtis J. Lile. Ed. D. Dissertation, Nova University. This work reviews testing workshops that were conducted for Fremont, California teachers, counselors, and administrators on test interpretation, administration, and analysis. A testing handbook, which includes reference materials designed for use in federally funded programs, and student diagnostic profiles, which aid in management of records is also provided.

ED 147 334 (1 mf)*

South Carolina Statewide Testing Program 1976-77. Summary Report, Office of Research Report Series, Vol. 1, No. 33. South Carolina State Dept. of Education, Columbia. Office of Research. This summary report explains the procedures used in the 1976-77 South Carolina Statewide Testing Program and the resulting data. The scores from the Comprehensive Tests of Basic Skills, Form S (CTBS/S)--which measure the reading, language, mathematics, science, social studies, and reference skills of students in grades 3, 6, and 11--are presented and are compared to previous years of testing and to national norms on the same tests.

ED 147 331 (4 mf)*

South Carolina Statewide Testing Program 1976-77. General Report, Office of Research Report Services. Vol. 1, No. 32. South Carolina State Dept. of Education, Columbia, Office of Research. This report presents the results of the 1976-77 South Carolina Comprehensive Tests of Basic Skills, Form S (CTBS/S) for grades 3, 6, and 11. It discusses the performance of the state's students in relation to the nation's performance, and analyzes the

results according to race, sex, Title I status, grade level, and skills tested. The purpose of the program, the methodology used, and the interpretation of the test scores are also described.

ED 147 326 (1 mf)*

District Utilization of the Statewide Testing Program. Virginia P. Rojas. Occasional Papers in Education. New Jersey State Dept. of Education, Trenton, Div. of Research, Planning, and Evaluation. This utilization survey tabulates the type of actions New Jersey school districts have taken upon being informed of the results of the EAP, or Educational Assessment Program--a statewide testing program used to assess students' reading and mathematics achievement. Over 72% of the responding districts used the statewide test data to modify their reading and mathematics curricula. It was also reported that 86.4% of the districts compared the test data from year to year; and that 36.8% used the data to initiate administrative changes involving the use of funds, the use or addition of staff, or the conduct of inservice workshops.

ED 116 263 (1 mf)*

Accountability and Testing. NAESP School Leadership Digest Second Series, Number 5. ERIC/CEM Research Analysis Series, Number 20. Jo Ann Mazzaella. (See page 68 for abstract) National Association of Elementary School Principals, Washington, D.C.; Oregon University, Eugene. ERIC Clearinghouse on Educational Management.

ED 109 246 (1 mf)*

The Appropriate and Inappropriate Uses of Grade Level Equivalents in School Evaluation. Walter E. Hathaway. Paper presented at the 1975 Meeting of the American Educational Research Association. This paper discusses the advantages, disadvantages, difference in variations, interpretations, interpolations and alternatives for using and reporting national grade level equivalent (GLE) scores and standardized tests of reading and mathematics for grades 4 and 8.

ED 099 425 (1 mf)*

Opening Institutional Ledger Books--A Challenge to Educational Leadership: Suggestions for Talking to School-Community Groups About Testing and Test Results. TM Report No. 28. Edwin P. Larsen. ERIC Clearinghouse on Tests, Measurement,

and Evaluation, Princeton, N.J. Sponsoring Agency: National Inst. of Education (DHEW), Washington, D.C. Three key areas are outlined dealing with the development of public understanding of testing: (1) Why tests are administered in schools: needs assessment, instructional program evaluation, materials selection, reporting to public, documenting individual growth, diagnostic analysis and planning, and instructional grouping. (2) Types of tests used, featuring explanations of achievement tests, Scholastic Aptitude Tests, interest tests, specialized aptitude tests, and personality tests. (3) Interpretation of test norms, raw scores, grade equivalent scores, percentile ranks and stanines, I.G. scores, and summarizing results (medians and quartiles). Methods used to chart test results of a school or district are discussed and suggestions made for the basic tools needed, the need for minimum use of numbers, and the facility of percentile ranks. Tables and charts for presenting statistical information are also proposed.

ED 095 212 (1 mf)*

Standardized Tests: A Handbook for Administration and Use.

Donald E. Lown. Greece Central School District, Rochester, N.Y. The purpose of this handbook is to provide information and guidelines for using data from standardized tests. The handbook explains the rationale for district-wide testing; discusses a district testing program; describes procedures for administering standardized tests; illustrates how to interpret and use standardized test data; and includes a glossary of terms.

*Indicates number of microfiche cards